Robust Recognition of Planar Mirrored Walls Using a Single View

Ali-akbar Agha-mohammadi and Dezhen Song

***Abstract*— We report a method for the detection and recognition of a large planar mirror based on the images captured by a monocular camera. We start with deriving a mirror transformation matrix in a homogeneous coordinate and geometric constraints for corresponding real and virtual feature points in the image. We find that existing feature detection methods are not reflection invariant. We introduce a secondary artificial reflection to virtual features to generate secondary features which are proven to share a rigid body motion relationship with the original feature set. We propose an iterative strategy to adjust the secondary mirror configuration so that existing feature matching methods can be used. The combined method yields a robust mirror detection algorithm which has been verified in physical experiments.**

I. Introduction

Mirrors are common objects in indoor environments and challenge robots in navigation. Cameras or light detection and ranging (LIDAR) cannot recognize mirrors because light simply bounces off the mirror surface. As service robots perform more and more tasks in indoor environments, the ability to recognize mirrors is necessary.

The ability of detecting a mirror or its own reflection in a mirror is a widely adopted test for intelligence levels [1]. Gallup first studies the self-directed behavior of animals using mirror-introduced tests [2]. A mirror and mark test is the frequently used method with the following setup: A subject has a mark that cannot be directly seen but is visible in the mirror. If there is increased exploration of the subject's own body and self-directed actions towards the mark, it implies that the subject recognizes the mirror image as self. Testing results from psychologists and biologists show that chimpanzees [2], dolphins [3], and magpies [4] have evident self-recognition in front of mirrors while gorillas do not.

If mirror or reflection detection is a challenging problem for mammals, there is no doubt that it is also a challenging recognition problem in robotic systems and computer vision. Oren and Nayar [5] analyze the characteristic and governing geometry of specular surfaces. However, their proposed method is limited to surfaces with high curvature and does not address detecting and modeling planar specular surfaces. Active lighting is also used in assisting the detection of specular surfaces. For example, Reiner and Donnaer [6] utilize stereo vision and a two dimensional array of light sources for constructing specular surfaces. Recently, Kutulakos and Steger [7] introduce a light-path triangulation method for constructing refractive and specular 3D objects, in which the light source must move along the light ray while the camera captures two consecutive images of the reflected light. Polarization imaging [8]–[11] is often used in mirror or reflective surface detection.

Following a minimalist's design, our robot only carries a single monocular camera and needs to recognize planes of large mirrored walls. We first derive a geometric constraint that relates the feature points of real objects to their reflections in the image, which are named as real-virtual pairs. We also find that existing advanced feature detection methods such as scale invariant feature transformation (SIFT) are not reflection invariant which leads to a high false negative rate in matching real-virtual pairs. To address the problem, we introduce a secondary artificial mirror reflection which converts virtual features into secondary features that share a rigid body motion relationship with the original features. The proposed method has been verified in physical experiments.

II. Problem Definition

We start with elaborating assumptions below:

1) The image taken by the robot contains a part of the planar mirror, some objects and their reflections.
2) All camera parameters are known.

There exist two coordinate frames in our system. Camera coordinate frame $\{C\}$ is a 3D right hand Cartesian coordinate system affixed to the camera with its Z-axis pointing out of the camera along the camera optical axis and its Y-axis pointing downward and being perpendicular to its Z axis. The origin of $\{C\}$ is the camera center $\widetilde{\mathbf{C}} = \mathbf{0}_{3\times 1}$. Frame $\{I\}$ is the 2D image coordinate, whose origin is the image of $\widetilde{\mathbf{C}}$, called principal point and is denoted by $\widetilde{\mathbf{c}}$.

The 3D points in $\{C\}$ and the 2D points in $\{I\}$ are denoted by a bold large-sized $\mathbf{X}$ and small-sized $\mathbf{x}$, respectively. $\widetilde{\mathbf{X}}$ and $\mathbf{X}$ are the inhomogeneous and the homogeneous forms of points, respectively. Moreover, if the homogeneous $\mathbf{X} = (\eta X, \eta Y, \eta Z, \eta)^T$, the relation between $\widetilde{\mathbf{X}}$ and $\mathbf{X}$ can be written as $\widetilde{\mathbf{X}} = (X, Y, Z)^T$ for $\eta \neq 0$.

Feature points are used for mirror detection:

1) Real feature points are the feature points corresponding to real objects in the environment.
2) Virtual feature points are the reflections of real feature points in the mirror.

If a real feature point in $\{C\}$ is $\mathbf{X}$, then the corresponding virtual feature point is denoted as $\mathbf{X}'$. Superscripts $'$ denotes the corresponding virtual feature points. The convention also applies to non-homogeneous points and 2D points.

With the above notions and assumptions defined, our planar mirror detection problem becomes

This work was supported in part by the National Science Foundation under CAREER grant IIS-0643298 and MRI-0923203.

A. Agha-mohammadi and D. Song are with CSE Department, Texas A&M University, College Station, TX 77843, USA, (email: aliagha@tamu.edu and dzsong@cs.tamu.edu)

Definition 1: Given images containing stationary objects and their reflections, recognize corresponding real-virtual pairs $\{\mathbf{x}_i, \mathbf{x}'_i\}$ in the image frame $\{I\}$ and estimate the mirror plane π_m in $\{C\}$.

To address this problem, we use a two-stage approach. First, we model the mirror reflection under camera perspective projection and derive the geometric constraints from known 2D real-virtual feature point pairs in $\{I\}$. Second, we present a robust estimation scheme to recognize 2D pairs from raw features. We begin with the first stage.

III. Modeling Mirror Reflection in the Image

To derive the geometric constraints for real-virtual feature pairs in $\{I\}$, we first analyze mirror reflection for a single real-virtual feature point pair and derive mirror normal.

A. Deriving a Minimal Solution for the Mirror Normal

Define the i-th real-virtual feature point pair as $\{\widetilde{\mathbf{X}}_i, \widetilde{\mathbf{X}}'_i\}$. The mirror plane must be the perpendicular bisector of the line segment connecting $\widetilde{\mathbf{X}}_i$ and $\widetilde{\mathbf{X}}'_i$. Three distinctive points, $\widetilde{\mathbf{X}}_i$, $\widetilde{\mathbf{X}}'_i$, and the camera center $\widetilde{\mathbf{C}}$ define a plane π_i. Therefore, the mirror plane has to be perpendicular to π_i. Back-projecting the corresponding 2D homogeneous real feature point, $\mathbf{x}_i = (\widetilde{\mathbf{x}}_i^T, 1)^T \in \{I\}$, leads to a line parameterized by λ [12], which contains all possible 3D points associated with $\mathbf{x}_i$,

$$\mathbf{X}_i(\lambda) = \mathrm{P}^{+}\mathbf{x}_i + \lambda\mathbf{C} \Rightarrow \widetilde{\mathbf{X}}_i(\lambda) = \frac{1}{\lambda_i}\mathrm{K}^{-1}\mathbf{x}_i, \qquad (1)$$

where P^{+} is the pseudo inverse of the camera perspective projection matrix, $\mathrm{P} = \mathrm{K}[\mathrm{I}_3|\mathbf{0}_{3\times1}]$, K is the intrinsic camera parameter matrix, I_3 is a 3×3 identity matrix, and $\mathbf{C} = (\widetilde{\mathbf{C}}^T, 1)^T$ is the homogeneous form of the camera center. Similarly, we can obtain $\widetilde{\mathbf{X}}'_i(\lambda')$ from the corresponding virtual feature point.

It is apparent that both $\widetilde{\mathbf{X}}_i(\lambda)$ and $\widetilde{\mathbf{X}}'_i(\lambda')$ are on π_i regardless of λ and λ'. Therefore, we can calculate $\mathbf{n}_{\pi_i}$, which is the normal of π_i,

$$\mathbf{n}_{\pi_i} = \widetilde{\mathbf{X}}_i \times \widetilde{\mathbf{X}}'_i, \qquad (2)$$

where '$\times$' is the cross product operator. With another pair of feature points, saying the j-th pair, we can define the plane π_j in a similar way. Since the mirror also has to be perpendicular to π_j, its normal can be computed as follows:

$$\mathbf{n}^{ij} = \mathbf{n}_{\pi_i} \times \mathbf{n}_{\pi_j} = \left(\widetilde{\mathbf{X}}_i \times \widetilde{\mathbf{X}}'_i\right) \times \left(\widetilde{\mathbf{X}}_j \times \widetilde{\mathbf{X}}'_j\right). \qquad (3)$$

Note that $\hat{\mathbf{n}}^{ij}$, the normalized version of $\mathbf{n}^{ij}$, does not depend on λ and λ' values. As a convention, a hat above any vector refers to the normalized form of that vector. For example, $\hat{\mathbf{n}} = \frac{\mathbf{n}}{\|\mathbf{n}\|}$ for vector $\mathbf{n}$. The first two entries of a vector is denoted by subscript $1:2$. For example, if $\mathbf{n} = [n_1, n_2, n_3]^T$, then $\mathbf{n}_{1:2} = [n_1, n_2]^T$.

Thus, (1-3) allow us to obtain mirror normal, $\hat{\mathbf{n}}^{ij}$, based on the i-th and the j-th feature point pairs. Note that the mirror normal vector $\hat{\mathbf{n}}$ is always pointing inside the mirror as shown in Fig. 1 in this paper.

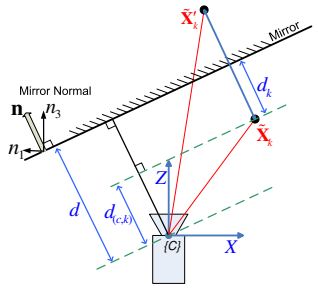

Fig. 1. A 2D view of the camera, the mirror, and the k-th pair of feature points. The view perspective is chosen so that the mirror degenerates to a line. The mirror is the perpendicular bisector of the line connecting the $\widetilde{\mathbf{X}}_k$ and $\widetilde{\mathbf{X}}'_k$. Two green dashed lines are parallel to the mirror plane.

B. Mirror Reflection Transformation

With the mirror normal ready, we can compute the transformation corresponding to the mirror reflection. Actually, the reflection transformation maps the real feature point $\widetilde{\mathbf{x}}_k$ to the corresponding virtual feature $\widetilde{\mathbf{x}}'_k$. The transformation is a function of mirror plane parameters. Define d as the distance from the camera center to the mirror. Denoting the distance from $\widetilde{\mathbf{X}}_k$ to the mirror by d_k and the distance from the camera center to $\widetilde{\mathbf{X}}_k$ along the mirror's normal by $d_{(c,k)}$, we can write down the following equations based on Fig. 1,

$$\begin{aligned} \widetilde{\mathbf{X}}'_k &= \widetilde{\mathbf{X}}_k + 2d_k\hat{\mathbf{n}}^{ij}, \\ d_k &= d - d_{(c,k)} = d - \widetilde{\mathbf{X}}_k^T\hat{\mathbf{n}}^{ij}. \end{aligned} \qquad (4)$$

Thus, we can write,

$$\widetilde{\mathbf{X}}'_k = \widetilde{\mathbf{X}}_k + 2(d - \widetilde{\mathbf{X}}_k^T\hat{\mathbf{n}}^{ij})\hat{\mathbf{n}}^{ij} = (\mathrm{I}_3 - 2\hat{\mathbf{n}}^{ij}\hat{\mathbf{n}}^{ij^T})\widetilde{\mathbf{X}}_k + 2d\hat{\mathbf{n}}^{ij}. \qquad (5)$$

Letting $\mathbf{X}_k = (\widetilde{\mathbf{X}}_k^T, 1)^T$ and $\mathbf{X}'_k = (\widetilde{\mathbf{X}}_k'^T, 1)^T$, we can write the affine transformation in (5) in the matrix form,

$$\mathbf{X}'_k = \mathbf{H}^{ij}\mathbf{X}_k, \qquad (6)$$

where $\mathbf{H}^{ij}$ is the reflection transformation matrix in $\{C\}$ which is determined by the i-th and j-th pairs of feature points,

$$\mathbf{H}^{ij} = \begin{pmatrix} \mathrm{I}_3 - 2\hat{\mathbf{n}}^{ij}\hat{\mathbf{n}}^{ij^T} & 2d\hat{\mathbf{n}}^{ij} \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix}. \qquad (7)$$

Substituting (1) into (6) leads to,

$$\mathbf{X}'_k = \mathbf{H}^{ij}\mathrm{P}^{+}\mathbf{x}_k + \lambda\mathbf{H}^{ij}\mathbf{C}. \qquad (8)$$

Projecting $\mathbf{X}'_k$ to the image plane results in the reflection transformation in $\{I\}$,

$$\mathbf{x}'_k = \mathrm{P}\mathbf{H}^{ij}\mathrm{P}^{+}\mathbf{x}_k + \lambda\mathrm{P}\mathbf{H}^{ij}\mathbf{C}. \qquad (9)$$

Note that depth d cannot be determined in a single view.

C. Constraints on Real-Virtual Feature Pairs

Eq. (9) provides a basis for finding the geometric constraint for real-virtual feature point pairs from a large number of noisy feature points in one image.

Lemma 1: If the feature points in i-th, j-th, and k-th pairs are matched correctly, then the following must be true,

$$\mathbf{x}_k'^T \mathcal{C}^{ij} \mathbf{x}_k = 0, \tag{10}$$

where $\mathcal{C}^{ij} = [\mathrm{K}\hat{\mathbf{n}}^{ij}]_\times (\mathrm{K}(\mathrm{I}_3 - 2\hat{\mathbf{n}}^{ij}\hat{\mathbf{n}}^{ij^T})\mathrm{K}^{-1})$ and matrix-vector multiplication format $[\mathbf{a}]_\times \mathbf{b}$ is used to represent the cross product of two vectors $\mathbf{a} \times \mathbf{b}$, where $[\mathbf{a}]_\times$ is a 3×3 skew-symmetric matrix composed by elements from vector $\mathbf{a} = (a_1, a_2, a_3)^T$ according to conventions in [12].

Proof: Eq. (9) introduces a line parameterized by λ, which contains two points $\mathrm{P}\mathbf{H}^{ij}\mathrm{P}^+\mathbf{x}_k$ and $\mathrm{P}\mathbf{H}^{ij}\mathbf{C}$ on the image plane. From the 2D projective geometry [12], the defined line by these two points can be written as follows:

$$l_k' = (\mathrm{P}\mathbf{H}^{ij}\mathbf{C}) \times (\mathrm{P}\mathbf{H}^{ij}\mathrm{P}^+\mathbf{x}_k). \tag{11}$$

Denoting $(\mathrm{I}_3 - 2\hat{\mathbf{n}}^{ij}\hat{\mathbf{n}}^{ij^T})$ by G^{ij}, we expand the above equation to,

$$l_k' = (2\mathrm{K}\hat{\mathbf{n}}^{ij}d) \times (\mathrm{K}\mathrm{G}^{ij}\mathrm{K}^{-1}\mathbf{x}_k) = 2d[\mathrm{K}\hat{\mathbf{n}}^{ij}]_\times (\mathrm{K}\mathrm{G}^{ij}\mathrm{K}^{-1})\mathbf{x}_k. \tag{12}$$

Defining $\mathcal{C}^{ij}$ as $[\mathrm{K}\hat{\mathbf{n}}^{ij}]_\times (\mathrm{K}\mathrm{G}^{ij}\mathrm{K}^{-1})$ and exploiting the fact that $\mathbf{x}_k'$ lies on l_k', we can write $\mathbf{x}_k'^T l_k' = 0$ and thus $\mathbf{x}_k'^T \mathcal{C}^{ij} \mathbf{x}_k = 0$. ■

The deviation of (10) from zero is often caused by the mismatch between feature points in the i-th, j-th, and k-th pairs. We denote this deviation by,

$$e_k^{ij} = \mathbf{x}_k'^T \mathcal{C}^{ij} \mathbf{x}_k = \mathbf{x}_k'^T l_k' = l_k \mathbf{x}_k^T. \tag{13}$$

where $l_k = \mathbf{x}_k'^T \mathcal{C}^{ij}$. This inspires a metric to measure the correspondence. We introduce D_k^{ij}, which depends on the distance from $\mathbf{x}_k'$ to l_k' and the distance from $\mathbf{x}_k$ to l_k,

$$(D_k^{ij})^2 = \Big(\frac{1}{l_{(k,1)}^2 + l_{(k,2)}^2} + \frac{1}{l_{(k,1)}'^2 + l_{(k,2)}'^2}\Big)(e_k^{ij})^2, \tag{14}$$

where $l_{(k,q)}$ and $l_{(k,q)}'$ are the q-th component of vectors l_k and l_k', respectively. D_k^{ij} can be viewed as a standardized version of e_k^{ij} with a clear geometric meaning.

IV. Robust Extraction of Feature Pairs

We now know that correctly-matched real-virtual pairs have to satisfy (10) which can be measured by D_k^{ij} in (14). With a set of raw feature points, we can revise random sample consensus (RANSAC) [13] framework to find the largest set of inliers which refer to pairs that conform to (10) and estimate mirror parameters. The remaining question is to choose the most appropriate feature extraction and matching methods for the mirror detection problem.

A. Limitations of Existing Feature Detection Methods

One natural choice is to apply the popular feature transformations such as scale invariant feature transformation (SIFT) [14] or its variations [15] to extract feature points from original pixel intensity data. Those feature points have been proven to be very robust in many applications. However, an immediate limitation appears when applying them to the mirror detection problem: *although SIFT features are purposefully designed to be scale invariant and even insensitive to some rotations, they are not reflection invariant.* The SIFT feature point vectors of a real-virtual pair do not necessarily match each other. This is true because a reflection mathematically cannot be represented as a combination of proper rotation and scaling operations.

One quick remedy to the problem is to reduce SIFT feature vectors from 128 dimensions to 2D position only to avoid the mismatch in part of SIFT vectors that describe neighboring characteristics with orientation information. We then apply RANSAC to see if we can find an inlier set that satisfies (10). We name this approach as the raw-SIFT approach. Unfortunately, this approach is very inefficient due to the small inlier ratio. Assume that there are a total number of m features which contain $q \ll m$ real-virtual potential feature pairs. Hence there are $\frac{q}{2}$ correct pairs needed to be found. We have a total of $\binom{m}{2} = \frac{m(m-1)}{2}$ pairs. Therefore, the inlier ratio is $\frac{q}{m(m-1)}$. Apparently, increasing the number of extracted features m actually decreases the ratio. A low inlier ratio means low signal to noise ratio and often leads to failure. This also explains why light weighted features such as Harris Corners [16] would not work well for this problem.

B. Converting Reflection to Rigid Body Motion

Therefore, we need to find a way to utilize the high dimensional SIFT feature vector to reduce the number of possible pairs to increase the inlier ratio. The intuition comes from a special case: Recall that $\mathbf{X}$ and $\mathbf{X}'$ refer to a real-virtual feature pair in $\{C\}$. Assume there exists a secondary mirror π_s sharing the same position and the opposite orientation of the unknown mirror π_m. Virtual feature point $\mathbf{X}'$ will have a secondary reflection point about π_s, which is defined as $\mathbf{X}''$. As a convention, we use $''$ to indicate feature points created by the secondary reflection. It is clear that $\mathbf{X}'' = \mathbf{X}$. Therefore, SIFT feature vectors of $\mathbf{X}$ and $\mathbf{X}''$ should match each other because their relationship is no longer a reflection. In fact, π_s does not need to be perfectly overlapped with π_m as we will show later. Introducing the artificial secondary reflection is the key to the problem. Even for an arbitrary π_s, we have the following observations:

Lemma 2: For any mirror pair π_m and π_s, the secondary feature point $\mathbf{X}''$ created from the reflection of $\mathbf{X}'$ about π_s can be obtained from the original feature point $\mathbf{X}$ by performing a rigid body motion (i.e. a combination of proper rotations and pure translations).

Proof: A rigid body motion can be represented by a rotation about and a translation along a screw axis [17]. The screw axis is defined by a unit vector $\mathbf{s}$ representing the screw direction. Point $\mathbf{s}_0$ lies on the screw axis and defines its original position. Therefore, the tuple $(\mathbf{s}, \mathbf{s}_0, \phi, t)$ describes a rigid body motion where ϕ is the rotation angle about the screw axis and t is the length of translation along the screw axis. The homogeneous transformation of the rigid body motion can be represented as:

$$\mathbf{A}(\mathbf{s}, \mathbf{s}_0, \phi, t) = \begin{pmatrix} \mathbf{R}_\phi^{\mathbf{s}} & \mathbf{q}_\phi^{\mathbf{s},\mathbf{s}_0} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}. \tag{15}$$

Based on [17] and simplifying the equation, we have,

$$\begin{aligned}\mathbf{R}_{\phi}^{\mathbf{s}} &= (\mathbf{s}\mathbf{s}^T - \mathrm{I}_3)(1-\cos\phi) + [\mathbf{s}]_\times \sin\phi + \mathrm{I}_3 \\ \mathbf{q}_{\phi}^{\mathbf{s},\mathbf{s}_0} &= t\mathbf{s} - (\mathbf{R}_{\phi}^{\mathbf{s}} - \mathrm{I}_3)\mathrm{diag}(\mathbf{s}_0). \end{aligned} \tag{16}$$

$\mathbf{R}_{\phi}^{\mathbf{s}}$ is a rotation matrix corresponding to a rotation of angle ϕ about the screw axis with direction $\mathbf{s}$ that goes through the origin, $\mathbf{q}_{2\theta}^{\mathbf{s},\mathbf{s}_0}$ is a translation vector, and $\mathrm{diag}(\mathbf{s}_0)$ refers to a diagonal matrix with its diagonal vector equal to $\mathbf{s}_0$. To prove Lemma 2, we need to show that

$$\mathbf{X}'' = \mathbf{A}\left(\mathbf{s}, \mathbf{s}_0, \phi, t\right)\mathbf{X}. \tag{17}$$

Define $\mathbf{H}$ and $\mathbf{H}''$ as homography matrices for the reflections with respect to π_m and π_s, respectively. Hence,

$$\mathbf{X}'' = \mathbf{H}''\mathbf{X}' = \mathbf{H}''\mathbf{H}\mathbf{X}. \tag{18}$$

Now we need to show if $\mathbf{H}''\mathbf{H}$ can be represented as $\mathbf{A}\left(\mathbf{s}, \mathbf{s}_0, \phi, t\right)$.

Define $\mathbf{n}$ as the normal of π_m and d as the depth which is the distance from the origin of $\{C\}$ to π_m. Similarly, we define normal $\mathbf{n}''$ and depth d'' for π_s. Based on (7), $(\mathbf{n}, d)$ and $(\mathbf{n}'', d'')$ determine $\mathbf{H}$ and $\mathbf{H}''$, respectively.

$$\mathbf{H} = \begin{pmatrix} \mathbf{G} & \mathbf{D} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix} \text{ and } \mathbf{H}'' = \begin{pmatrix} \mathbf{G}'' & \mathbf{D}'' \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}, \tag{19}$$

where $\mathbf{G} = \mathrm{I} - 2\hat{\mathbf{n}}\hat{\mathbf{n}}^T$, $\mathbf{D} = 2\zeta d\hat{\mathbf{n}}$, $\mathbf{G}'' = \mathrm{I} - 2\hat{\mathbf{n}}''\hat{\mathbf{n}}''^T$, and $\mathbf{D}'' = 2\zeta''d''\hat{\mathbf{n}}''$. Sign variable ζ equals to $+1$ or -1 if the camera projection center is in front of or behind mirror π_m, respectively. Similarly, sign variable ζ'' equals to $+1$ or -1 if the camera projection center is in front of or behind the secondary mirror π_s, respectively. Multiplying these two homography matrices, we have:

$$\mathbf{H}''\mathbf{H} = \begin{pmatrix} \mathbf{R} & \mathbf{q} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}, \tag{20}$$

where $\mathbf{R} = \mathbf{G}''\mathbf{G}$ and $\mathbf{q} = \mathbf{G}''\mathbf{D} + \mathbf{D}''$.

Since $\mathbf{G} = \mathrm{I} - 2\hat{\mathbf{n}}\hat{\mathbf{n}}^T$ and $\mathbf{G}'' = \mathrm{I} - 2\hat{\mathbf{n}}''\hat{\mathbf{n}}''^T$, matrices $\mathbf{G}$ and $\mathbf{G}''$ are symmetric with determinants of -1. Consequently, we have

$$\mathbf{R}\mathbf{R}^T = \left(\mathbf{G}''\mathbf{G}\right)\left(\mathbf{G}''\mathbf{G}\right)^T = \mathbf{G}''\mathbf{G}\mathbf{G}^T\mathbf{G}''^T = \mathrm{I}, \tag{21}$$

$$\det(\mathbf{R}) = \det(\mathbf{G}'')\det(\mathbf{G}) = (-1)(-1) = 1. \tag{22}$$

Since $\mathbf{R}$ is an orthogonal matrix with a determinant of 1, $\mathbf{R}$ must be a proper rotation matrix. According to [17] and plugging in $\mathbf{H}''\mathbf{H}$ from (20), ϕ, $\mathbf{s}$, and t can be obtained,

$$\phi = 2\theta, \quad \mathbf{s} = \frac{(\hat{\mathbf{n}} \times \hat{\mathbf{n}}'')}{\sin(\theta)}, \quad t = 0. \tag{23}$$

The result in (23) is true for non-parallel mirrors. If the mirrors are parallel, matrix $\mathbf{A}$ degenerates to a pure translation and $\hat{\mathbf{n}}'' = \hat{\mathbf{n}}$, the results are $\phi = 0$, $\mathbf{s} = \hat{\mathbf{n}}$, $t = 2(\zeta''d'' - \zeta d)$. The same conclusion holds for the lemma. ■

Lemma 2 shows that two consecutive reflections are equivalent to a rigid body motion. We are one step closer to utilize SIFT feature vectors. However, performing such a 3D reflection is not straightforward because we do not have 3D positions of features. Fortunately, we can adjust the position and orientation of π_s to transfer the secondary 3D reflection to a 2D reflection in $\{I\}$ about a line.

C. Reducing Secondary Reflection from 3D to 2D

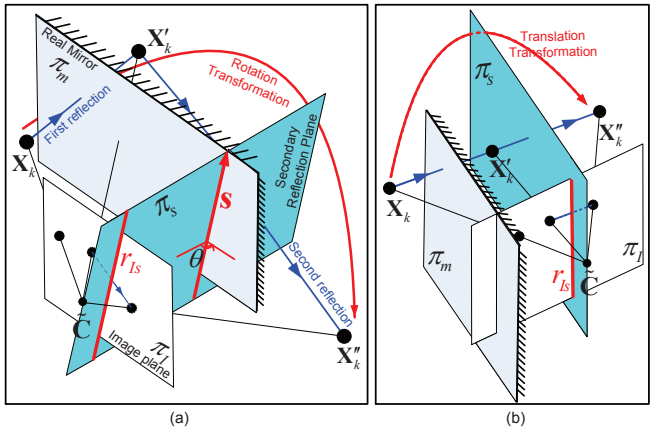

Fig. 2. (a) The configuration of π_I, π_s, and π_m. π_s is perpendicular to π_I and goes through $\widetilde{\mathbf{C}}$. (b) The ideal configuration when $\pi_s \parallel \pi_m$ and $\theta = 0$.

Denoting the image plane as π_I, the key of the reflection reduction is to force the secondary reflection plane π_s to be perpendicular to π_I and to pass through camera projection center $\widetilde{\mathbf{C}}$. Fig. 2(a) illustrates the plane relationship. Recall that $\mathbf{n}''$ is the normal of π_s. π_s is not unique because a different $\mathbf{n}''_{1:2}$ would result in a different π_s. The intersection between π_s and π_I projects itself to a line r_{Is} on π_I which goes through the principal point. Also, the normal of r_{Is} equals to $\mathbf{n}''_{1:2}$. Since $\widetilde{\mathbf{C}}$ lies on π_s, $\mathbf{D}'' = \mathbf{0}_{3\times 1}$ must be true according to (19). The conditions of $\pi_s \perp \pi_I$ and $\mathbf{D}'' = \mathbf{0}_{3\times 1}$ lead to the fact that the 3D reflection about π_s is equivalent to a 2D reflection about r_{Is} in π_I based on (9). Recall that $\mathbf{x}''_k$ is the imaging point of $\mathbf{X}''_k$, we know

$$\mathbf{x}''_k = \mathbf{H}''_{2D}\mathbf{x}'_k, \text{ where } \mathbf{H}''_{2D} = \begin{pmatrix} \mathrm{I}_2 - 2\hat{\mathbf{n}}''_{1:2}\hat{\mathbf{n}}''^T_{1:2} & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & 1 \end{pmatrix}, \tag{24}$$

and

$$\mathbf{x}''_k = \mathbf{H}''_{2D}\mathrm{P}\mathbf{H}\mathrm{P}^{+}\mathbf{x}_k + \lambda\mathbf{H}''_{2D}\mathrm{P}\mathbf{H}\mathbf{C}. \tag{25}$$

To find the matching features, we reflect the whole image I about r_{Is}. Then applying SIFT, we extract features from original image I and the reflected image I''. Denote these two sets of features by $\mathscr{I} = \{\mathbf{x}_s\}_{s=1}^{m}$ and $\mathscr{I}'' = \{\mathbf{x}''_s\}_{s=1}^{m''}$, respectively. Performing matching between two sets of features $\mathscr{I}$ and $\mathscr{I}''$ results in a set of w matched pairs $\mathscr{M}'' = \{(\mathbf{x}_s, \mathbf{x}''_s)\}_{s=1}^{w}$. Since $\mathbf{H}''_{2D}$ and d'' are known, (24) introduces a known one-to-one mapping and we can retrieve $\widetilde{\mathbf{x}}'_k$'s from $\widetilde{\mathbf{x}}''_k$'s that make a set $\mathscr{I}' = \{\mathbf{x}'_s\}_{s=1}^{w}$. Substituting corresponding $\widetilde{\mathbf{x}}''_k$'s and $\widetilde{\mathbf{x}}'_k$'s in $\mathscr{M}''$, we end up with $\mathscr{M} = \{(\mathbf{x}_s, \mathbf{x}'_s)\}_{s=1}^{w}$.

D. Adjusting Image Plane to Reduce Perspective Changes

According to [14], SIFT is designed to be scale invariant and also works very well if the perspective change caused by a rotation is no more than 40 degrees. An arbitrary rigid body motion may include a rotation over the 40-degree limit and decrease the effectiveness of SIFT feature matching. To

ensure the quality of SIFT matching, it is desirable if the angle between π_s and π_m, $\theta = \measuredangle(\pi_s, \pi_m) \leq 20^\circ$ since the reflection doubles the angle. To achieve this, we design an iterative procedure.

Initially, r_{Is} is chosen randomly and does not change during the computation. Hence the relative position and orientation between π_s and π_I is fixed at all time. Any rotation applied to π_I is also applied to π_s. Knowing that the matching error will be ignorable when the condition $\theta < 20^\circ$ is satisfied, we rotate π_I to search for this configuration. The ideal case happens when $\theta = 0$ (Fig. 2(b)). This means that feature vectors corresponding to $\mathbf{X}$ and $\mathbf{X}''$ only have scale difference, which is perfect for SIFT matching.

The initial run with the randomly selected r_{Is} and the corresponding π_s might not satisfy the condition $\theta \leq 20^\circ$. To refine the solution, we rotate π_I to repeat the whole procedure to refine the solution. Each rotation of π_I is done by applying a homography H_R that is created from a two-step rotation. First, we rotate the π_I about the camera Y axis by angle α to make it perpendicular to π_m. α is the angle between π_m and the optical axis measured on the $X - Z$ plane of camera coordinates. Note that this rotation is a standard Y-axis rotation that can be represented by matrix R^y_α. Then rotated mirror normal is $\mathbf{n}_R = R^y_\alpha \mathbf{n}$. Second, we rotate π_I using R^z_β about z axis by angle β to make π_m and π_s parallel. The normal of intersecting line of the rotated mirror plane with image plane is $\mathbf{n}_{R_{1:2}}$. Thus, β is the angle between $\mathbf{n}_{1:2}$ and $\mathbf{n}_{R_{1:2}}$ on π_I. α and β are computed as follows:

$$\alpha = \tan^{-1}\left(\frac{n_1}{n_3}\right), \quad \beta = \cos^{-1}\left(\frac{\mathbf{n}^T_{R_{1:2}}\mathbf{n}''_{1:2}}{\|\mathbf{n}_{R_{1:2}}\|\|\mathbf{n}''_{1:2}\|}\right). \quad (26)$$

All above procedure is projectively equivalent to applying the following homography to the image,

$$H_R = K\left(R^z_\beta R^y_\alpha\right)K^{-1}. \quad (27)$$

After each rotation, we remove self-matching pairs and reject outliers from $\mathscr{M}$ using RANSAC framework. Define $\mathscr{C}^*$ as the largest inlier pair set. Finally, Maximum Likelihood Estimator (MLE) is used to compute the best mirror normal from the matches in $\mathscr{C}^*$. The overall robust mirror detection algorithm (RMDA) is recapped in Algorithm 1, in which t_h is a threshold for distinguishing inlier and outlier pairs. Iteration number N can be determined by choosing 95% probability of finding the solution in RANSAC. Note that since we only need $\theta \leq 20^\circ$ instead of $\theta = 0$, the outer loop should converge within 3 iterations with each iteration to refine the solution. If the loop cannot converge, it usually means the assumptions in Section II are violated and the method fails.

Algorithm 1: Robust Mirror Detection Algorithm

input : Original captured image I
output : Mirror normal $\mathbf{n}$ and the inlier feature set $\mathscr{C}^*$
Choose a random r_{Is} passing the principal point of $\{I\}$;
for $counter = 1$ **to** 3 **do**
  Reflect I about r_{Is} using (24) to get I'';
  Extract SIFT feature sets $\mathscr{I} = \{\mathbf{x}_s\}_{s=1}^{m}$ from I and $\mathscr{I}'' = \{\mathbf{x}''_s\}_{s=1}^{m''}$ from I'';
  Perform SIFT descriptor matching using [14] between $\mathscr{I}$ and $\mathscr{I}''$ to obtain $\mathscr{M}'' = \{(\mathbf{x}_s, \mathbf{x}''_s)\}_{s=1}^{w}$;
  Compute $\mathscr{M} = \{(\mathbf{x}_s, \mathbf{x}'_s)\}_{s=1}^{w}$ using inverse of (24);
  for $u = 1$ **to** N **do**
    Randomly sample 2 pairs from $\mathscr{M}$ and denote them by $S_u = \{(\mathbf{x}_i, \mathbf{x}'_i), (\mathbf{x}_j, \mathbf{x}'_j)\}$;
    Compute $\mathbf{n}^{ij}$ based on (3);
    $\mathscr{C}_u = S_u$; $k = 2$;
    for $a = 1$ **to** w **do**
      pair $k = \mathscr{M}[a]$;
      Compute D_k^{ij} using (14);
      if $|D_k^{ij}| < t_h$ **then**
        $\mathscr{C}_u = \mathscr{C}_u \cup \{\text{pair } k\}$;
        $k = k + 1$;
  $u_{\max} = \arg\max_u |\mathscr{C}_u|$;
  $\mathscr{C}^* = \mathscr{C}_{u_{\max}}$;
  Apply MLE on $\mathscr{C}^*$ to obtain $\mathbf{n}$;
  Compute α and β using (26);
  Rotate I using H_R from (27);
  if $\arccos(\widehat{\mathbf{n}}^T\widehat{\mathbf{n}}'') \leq 20^\circ$ **then**
    return $\mathbf{n}$;
return failure;

V. Experiments

We have implemented our mirror detection algorithm using Matlab on a PC laptop with a Windows XP operating system. For the SIFT algorithm, we have used an open source implementation of SIFT in [18]. The images have a resolution of 640×480 pixels. The testing images are taken using a pre-calibrated Canon A1000 digital camera.

In the first experiment, we illustrate the geometric constraint in Lemma 1 using a sample case (Fig. 3(a)). The geometric constraint in Lemma 1 can be visualized as a point-line relationship as shown in (13). For a real feature point $\mathbf{x}_k$, we can obtain a line l'_k using (12). According to Lemma 1, the corresponding virtual feature point $\mathbf{x}'_k$ must lie on l'_k. Fig. 3(a) shows this is true for correctly matched pairs. To avoid a cluttered figure, we only show 10 matched pairs (totally 20 feature points) in Fig. 3(a).

It is worth noting that the seventh pair is not a true pair because it is not resulted from matching a real-virtual feature pair. The false matching is due to the existing symmetry in scene. Most of such false detections have been removed using RANSAC. This particular spurious match could not be removed because the direction of symmetry happens to be the same as mirror normal. However, such case can be easily handled using depth information.

In the second experiment, we compare the raw-SIFT approach to our RMDA. We have tested both approaches on 51 images taken from large mirrors in gymnasiums, shopping malls, showrooms, and etc. These photos contain different mirror normal directions and different mirror sizes with different scenes. Fig. 3(b) shows some of these scenes. For each image, we manually select correctly matched real-virtual pairs to compute mirror normal as a ground truth. If RMDA finishes successfully, and the angle between the mirror normal from RMDA and ground truth mirror normal is less than 5 degrees, the method succeeds in recognizing

(a) (b) (c)

Fig. 3. (a) An illustration of feature points (small yellow dots) and geometric constraints (pink lines). Units are pixels. The number on each feature point shows that to which pair it belongs. (b) Sample test images. (c) A comparison of successful rates of RMDA (shown in the solid line) and the raw-SIFT method (shown in the dashed line).

the mirror. The performance of the algorithms depends on the number of correctly-matched features. Adjusting the SIFT strength threshold changes the number of features and affect the inlier ratio. A smaller strength threshold yields a larger number of feature points. Fig. 3(c) shows average success rates over all testing images for different values of the strength threshold. To help understanding how the strength threshold affects the number of feature points, the top horizontal axis of Fig. 3(c) provides the number of features for the corresponding threshold for a sample image, which is the bottom-left image in Fig. 3(b). As shown in Fig. 3(c), success rates of RMDA steadily increases as the number of features increases while the raw-SIFT cannot utilize the increased features. For failure cases, we find that lack of features is the primary reason.

VI. Conclusion and Future Work

We proposed a mirror detection method for recognizing a large planar mirror using an image captured by a monocular camera mounted on a mobile robot. We derived a closed form solution for computing the mirror normal and a geometric constraint between feature point pairs. We found that existing advanced feature detection methods are not reflection invariant. We introduced an artificial secondary mirror into the system to transform the reflection relationship to an orientation preserving transformation. We also design an iterative method to adjust the configuration of the second mirror to enable SIFT descriptor matching. Combining the results, we proposed a robust mirror detection algorithm and the experimental results confirmed our analysis. In the future, we will work on depth assisted mirror detection and mirror boundary segmentation problems.

Acknowledgement

We thank J. Zhang and C. Kim for their insightful inputs and help with the experiments. We thank Y. Xu, W. Li, Y. Lu, and H. Li for their feedback.

References

[1] W. G. Walter, "An imitation of life," *Scientific American*, vol. 182, no. 2, pp. 42–45, 1950.

[2] G. Gallup, "Chimpanzees: Self-recognition," *Science*, vol. 167, no. 3914, pp. 86–87, January 1970.

[3] D. Reiss and L. Marino, "Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence," *PNAS*, vol. 98, no. 10, pp. 5937–5942, May 2001.

[4] H. Prior, A. Schwarz, and O. Gntrkn, "Mirror-induced behavior in the magpie (pica pica): Evidence of self-recognition," *PLoS Biol*, vol. 6, no. 8, p. e202, 08 2008.

[5] M. Oren and S. K. Nayar, "A theory of specular surface geometry," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 105–124, September 1997.

[6] K. Reiner and K. Donner, "Stereo vision on specular surface," in *Proceedings of 4th IASTED International Conference on Visualization, Imaging and Image processing, Marbella, Spain*, September 2004.

[7] K. N. Kutulakos and E. Steger, "A theory of refractive and specular 3d shape by light-path triangulation," *International Journal of Computer Vision*, vol. 76, no. 1, pp. 13–29, January 2008.

[8] M. Ferraton, C. Stolz, and F. Meriaudeau, "surface reconstruction of transparent objects by polarization imaging," in *IEEE International Conference on Signal Image Technology and Internet Based Systems*, 2008.

[9] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 73–82, January 2004.

[10] S. Rahmann and N. Canterakis, "Reconstruction of specular surfaces using polarization imaging," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.

[11] O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria, "Active lighting applied to 3d reconstruction of specular metallic surfaces by polarization imaging," *Applied Optics*, vol. 45, pp. 4062–4068, January 2006.

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision, 2nd Edition*. Cambridge University Press, 2004.

[13] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, June 1981.

[14] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 4, pp. 91–110, Nov. 2004.

[15] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *9th European Conference on Computer Vision (ECCV)*, Graz, Austria, May 2006, pp. 404–417.

[16] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*, vol. 15, 1988, pp. 147–151.

[17] L.-W. Tsai, *Robot Analysis: The Maechanics of Serial and Parallel Manipulators*. John Wiley and Sons, Inc, 1999.

[18] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.